

LOCALLY ADAPTIVE GREEDY APPROXIMATIONS FOR ANISOTROPIC PARAMETER REDUCED BASIS SPACES

YVON MADAY AND BENJAMIN STAMM

ABSTRACT. Reduced order models, in particular the reduced basis method, rely on empirically built and problem dependent basis functions that are constructed during an off-line stage. In the on-line stage, the precomputed problem-dependent solution space, that is spanned by the basis functions, can then be used in order to reduce the size of the computational problem. For complex problems, the number of basis functions required to guarantee a certain error tolerance can become too large in order to benefit computationally from the model reduction. To overcome this, the present work introduces a framework where local approximation spaces (in parameter space) are used to define the reduced order approximation in order to have explicit control over the on-line cost. This approach also adapts the local approximation spaces to local anisotropic behavior in the parameter space. We present the algorithm and numerous numerical tests.

1. INTRODUCTION

The recent progresses in the numerical simulation of physical phenomena obtained through the combination of robust and accurate approximation methods and faster and larger computational platforms have permitted to investigate more and more complex problems with improved reliability. These progresses have, in turn, lead to new demands for the numerical simulations that are not only used to understand a given state but investigate control and optimization problems. In such problems, a large number of similar computations of a parameter-dependent model have to be performed associated with different values of the parameter chosen either randomly or guided by some well suited recursive algorithm.

Faster solution algorithms are often not sufficient to achieve these new demands in many engineering applications and reduced numerical methods have been proposed as surrogates to standard numerical high fidelity approximations of given mathematical models. Among these methods, the reduced basis (RB) methods [16, 14, 15] is a class of higher-order mathematical technique that has been widely developed and gained in generality and reliability during the last decade. The greedy algorithm [18] plays a crucial ingredient and will be the main focus of the present paper. For parametrized time-dependent problems, the RB (greedy-based) strategy can be successfully combined with the proper orthogonal decomposition (POD) [10] as is illustrated in [8, 12].

The basic idea behind these reduced order methods is the notion of small Kolmogorov n -width of the set of all solutions under variation of the parameter, i.e. that the solution manifold, embedded in a Hilbert space \mathbb{W} , can be well approximated with a well chosen n -dimensional subspace of \mathbb{W} . In essence, RB methods

Key words and phrases. model reduction, reduced basis method, greedy algorithm .

are based on a two steps strategy : the first step (off-line stage) allows to select particular instances of the parameters, for which a very accurate approximation of the solution is computed : those solutions constitute the reduced basis. In a second step (the on-line stage), the generic solutions (for other instances of the parameter) are approximated by a linear combination of these reduced basis functions. The interest of the approach lies in the fact that, in cases where the Kolmogorov n -width is small — and there are many indications for such a smallness: (i) a priori, such as regularity of the solutions as a function of the parameters when the set of parameters lies in a finite dimensional space, and (ii) a posteriori, revealed by a large number of simulation — these reduced model approximations require very few degrees of freedom and in some sense are close in spirit to other high order approximation methods like spectral methods, but with an improved efficiency for what concerns the computational complexity.

It is well known that high order methods generally take advantage of a global approach by using basis functions that have a large support that, combined with higher regularity of the solution to be approximated (going some times up to analyticity) allows to get, with very few degrees of freedom, a very good accuracy. For most practical design problems in engineering though, the solution is not analytical and most of the time regularity exists but only locally which precludes the interest of global approximations. This is the reason why, for instance, by breaking the global framework to locally piecewise global approaches, the spectral element method reveals superiority with respect to the plain spectral method: a trade off between locality and globality is generally preferred as is demonstrated in e.g. [2] for approximation in spatial direction by spectral element approximations. The same causes imply the same effects in the parameter directions for RB approximations, this is the reason why recently, some works have been devoted to proposing ways of adding locality to the reduced methods yielding parameter subdomain domain refinement.

In this paper we shall focus on the certified RB framework for which, the construction of basis functions during the first stage of the algorithm is, as in most of the current approaches, performed through a greedy strategy based on an a posteriori error estimator. For either high dimensional parameter spaces or spaces with large measure one may encounter the problem that the size of the reduced basis turns out to be larger than desired. Following the lines drawn above, first ideas in this context has been presented by Eftang, Patera and Rønquist [4] and further in [5, 6, 7] where the parameter space is decomposed into cells where different reduced basis sets are assembled. This approach presents clear advantages in the size of the matricial system that appears in the on-line solution procedure, and corroborates the natural feeling that, in order to approximate the solution at a given parameter, primarily those solutions in the reduced basis corresponding to parameters that are close to the parameter we are interested in are to be involved in the linear approximation.

A drawback of the current approach [4, 5, 6, 7] however is, that in two adjacent parameter-subdomains, some of the parameters that are selected may be very close or even lie both on the common interface. This is a situation that may be encountered very often when dealing with high dimensional parameter spaces, and in addition it is known from computations that the greedy algorithm has the tendency to pick sample points lying on the boundary of the parameter space. In this case,

the approximation on one of the subdomains does not benefit of the computation that was done for the other subdomain. Parameter domain decomposition may thus not be the ultimate approach. Another drawback of all classical greedy approaches is to be unable to master the size of the reduced basis and in consequence the size of the discrete system that will be solved in the on-line procedure.

These reflections have motivated us to investigate the alternative discussed in this paper which introduces a completely new framework, which should be seen as a first layout of our ideas where it is unfortunately impossible to answer all interesting questions related to this approach without losing the big picture.

The outline of the paper is as follows. In Section 2 we present the standard greedy algorithm that is widely used in reduced basis computations in order to have a common ground for the further discussion. Then, we introduce in Section 3 the concept of local approximation spaces, that is, if a reduced order approximation needs to be computed at a parameter point $\boldsymbol{\mu}$ the approximation space is spanned by the N closest precomputed basis functions. Thus, the construction of the local approximation spaces is based on a distance function which is constructed in Section 4 in an empirical manner taking into account the geometry of the manifold of the solution set including the local anisotropies in the parameter space. Section 5 introduces the concept of a varying training set that is chosen in an optimal fashion using the possibly anisotropic distance function if the computing resources are limited to handle a given (and too small) number of training points. In Section 6 we explain some practical aspects of the on-line implementation, in particular how the different snapshots can be orthonormalized at the on-line stage. In Section 8 we present some numerical examples which uses the best approximation (L^2 -projection) for an explicitly given family of functions whereas an example using a reduced basis framework is given in Section 9. Finally, in Section 10 we draw the conclusions.

2. THE GREEDY ALGORITHM FOR REDUCED BASIS APPROXIMATIONS

Let us first introduce a classical greedy algorithm to have a common ground to present our ideas. In a general setting, it is assumed that for each parameter value $\boldsymbol{\mu}$ in a parameter domain $\mathbb{P} \subset \mathbb{R}^p$, a $\boldsymbol{\mu}$ -dependent function $v(\boldsymbol{\mu}) \in \mathbb{W}$ of the variable \boldsymbol{x} in a bounded domain $\Omega \subset \mathbb{R}^d$ (e.g. $d = 2$ or $d = 3$) can be computed. The space \mathbb{W} denotes some functional space of functions defined on Ω and we denote the solution manifold by

$$\mathbb{W}_{\mathbb{P}} = \{v(\boldsymbol{\mu}) \in \mathbb{W} \mid \boldsymbol{\mu} \in \mathbb{P}\}.$$

In order to fix the ideas this can be that $v(\boldsymbol{\mu})$ is a solution to a parameter dependent partial differential equation, an approximation of which can be computed by the e.g. classical finite element or spectral method. Let

$$\mathbb{S}_N = \{\boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^N\}$$

be a collection of N parameters in \mathbb{P} and

$$\mathbb{W}_N = \text{span}\{v(\boldsymbol{\mu}^1), \dots, v(\boldsymbol{\mu}^N)\}$$

be the approximation space associated to the set \mathbb{S}_N . A projection operator $P_N : \mathbb{W} \rightarrow \mathbb{W}_N$ is supposed to exist that can be either a parameter dependent Ritz-projection (reduced basis methods in case of a parameter dependent PDE) [14, 17] or simply a L^2 -projection. Further, it is assumed that for each parameter value

$\boldsymbol{\mu} \in \mathbb{P}$ an error estimator $\eta(\boldsymbol{\mu}; \mathbb{W}_N)$ of the approximation of $v(\boldsymbol{\mu})$ by $P_N(v(\boldsymbol{\mu}))$ can be computed (this can be e.g. through a posteriori analysis of the residual, see eg [14]). We assume that there exists two constants $c > 0$ and $C > 0$ such that

$$c\eta(\boldsymbol{\mu}; \mathbb{W}_N) \leq \|v(\boldsymbol{\mu}) - P_N(v(\boldsymbol{\mu}))\| \leq C\eta(\boldsymbol{\mu}; \mathbb{W}_N),$$

for an appropriate norm on \mathbb{W} . The case where $\eta(\boldsymbol{\mu}; \mathbb{W}_N)$ describes the exact error is thus not excluded in this general framework.

The following sketch represent a typical greedy algorithm: Let Ξ_{train} be a chosen finite training set $\Xi_{\text{train}} \subset \mathbb{P}$ of representative points.

ClassicalGreedy

Input: tol.

Output: $N, \mathbb{W}_N, \mathbb{S}_N$.

1. Choose (possibly randomly) $\boldsymbol{\mu}^1 \in \Xi_{\text{train}}$, set $\mathbb{S}_1 = \{\boldsymbol{\mu}^1\}$, $\mathbb{W}_1 = \text{span}\{v(\boldsymbol{\mu}^1)\}$ and $N = 1$, $\text{err} = \max_{\boldsymbol{\mu} \in \Xi_{\text{train}}} \eta(\boldsymbol{\mu}; \mathbb{W}_1)$.
2. While $\text{err} > \text{tol}$
3. Find $\boldsymbol{\mu}^{N+1} = \arg\max_{\boldsymbol{\mu} \in \Xi_{\text{train}}} \eta(\boldsymbol{\mu}; \mathbb{W}_N)$, $\text{err} = \max_{\boldsymbol{\mu} \in \Xi_{\text{train}}} \eta(\boldsymbol{\mu}; \mathbb{W}_N)$.
4. Compute $v(\boldsymbol{\mu}^{N+1})$, set $\mathbb{S}_{N+1} = \mathbb{S}_N \cup \{\boldsymbol{\mu}^{N+1}\}$ and set $\mathbb{W}_{N+1} = \text{span}\{\mathbb{W}_N, v(\boldsymbol{\mu}^{N+1})\}$.
5. Set $N := N + 1$.
6. End while.

Algorithm 1: Classical greedy algorithm.

Remark 2.1. *The introduction of the finite set Ξ_{train} is due to practical implementation, because the entire scan of \mathbb{P} is impossible: Ξ_{train} has to be finite, not too large in order the previous algorithm is not too lengthy, not too small in order to represent well \mathbb{P} . Actually the definition of this finite set can evolve during the algorithm, see [7]. We shall elaborate on this in the framework of our approach later in the paper (Section 5). An alternative approach is presented in [1], where finding the maximum is stated as a constraint optimization problem.*

Let \tilde{N} denote the final size of the reduced basis space $\mathbb{W}_{\tilde{N}}$ such that the accuracy criterion

$$\max_{\boldsymbol{\mu} \in \Xi_{\text{train}}} \eta(\boldsymbol{\mu}; \mathbb{W}_{\tilde{N}}) \leq \text{tol}$$

is satisfied. Note that this final size is a consequence of the tolerance that has been chosen, hence, for a prescribed tolerance, we have not the hand on the complexity of the RB approximation method.

3. LOCALLY ADAPTIVE GREEDY APPROXIMATIONS

The new approach pursued in this work is in the same spirit of partitioning the parameter set as in [4, 5, 6, 7]. However, the difference is listed below:

- [1] We do not impose a clear partition of the parameter space but rather collect a global set of K sample points \mathbb{S}_K , they are preliminary selected in the *off-line* stage and the corresponding basis functions are computed and stored. This step is elaborated in Section 3.2.

- [2] We choose a priori the size N ($\leq K$) of the system we want to solve *on-line*. Thus, the resulting complexity of the *on-line* stage is user controlled.
- [3] When a reduced basis approximation is to be computed for a certain given parameter value $\boldsymbol{\mu} \in \mathbb{P}$ in the *on-line* stage, we only use the N precomputed basis functions whose corresponding parameter values lie in a ball around $\boldsymbol{\mu} \in \mathbb{P}$. This step is elaborated in Section 3.1.
- [4] The distance function used to define the ball, containing the N closest basis functions, is an empirically constructed distance function that is problem dependent and can reflect anisotropies in parameter space and its local changes. The construction of the distance function is presented in Section 4.

3.1. Local approximations spaces. As already anticipated, for a certain given parameter value $\boldsymbol{\mu} \in \mathbb{P}$ in the *on-line* stage we only use the N basis functions whose parameter values lie in a ball

$$(1) \quad B_{\boldsymbol{\mu}} = \{\tilde{\boldsymbol{\mu}} \in \mathbb{P} \mid d(\boldsymbol{\mu}, \tilde{\boldsymbol{\mu}}) \leq r(\boldsymbol{\mu})\},$$

for a given distance function $d(\cdot, \cdot)$. The radius $r(\boldsymbol{\mu})$ is tuned in such a way that there are actually N basis functions in the ball. Denoting by \mathbb{S}_K the set of sample points for which the corresponding functions are precomputed at the off-line stage, the local sample space is defined by

$$\mathbb{S}_{\boldsymbol{\mu}} = B_{\boldsymbol{\mu}} \cap \mathbb{S}_K = \{\tilde{\boldsymbol{\mu}} \in \mathbb{S}_K \mid \tilde{\boldsymbol{\mu}} \in B_{\boldsymbol{\mu}}\}$$

with cardinality equal to N^1 . Further the local reduced basis approximation space shall be defined by $\mathbb{W}_{\boldsymbol{\mu}} = \text{span}\{v(\tilde{\boldsymbol{\mu}}) \mid \tilde{\boldsymbol{\mu}} \in \mathbb{S}_{\boldsymbol{\mu}}\}$ and its associated local projection by $P_{\boldsymbol{\mu}} : \mathbb{W} \rightarrow \mathbb{W}_{\boldsymbol{\mu}}$ in order to define the reduced basis approximation $P_{\boldsymbol{\mu}}(v(\boldsymbol{\mu}))$ to $v(\boldsymbol{\mu})$. Also in this local framework, the Galerkin setting can be used to design a posteriori estimates to quantify the error between $v(\boldsymbol{\mu})$ and $P_{\boldsymbol{\mu}}(v(\boldsymbol{\mu}))$, that we shall denote by $\eta(\boldsymbol{\mu}; \mathbb{W}_{\boldsymbol{\mu}})$.

A similar approach is presented in [11] where the snapshots are chosen also in the on-line stage but such that the approximation error is maximally reduced. Thus, in general, no local neighborhood is required.

One basic ingredient of this local reduced basis approximation is thus the distance function $d(\cdot, \cdot)$; the radius is adjusted to select the total number of modes N for each value of $\boldsymbol{\mu}$ whereas the distance function can be chosen isotropic or better accounts for anisotropies in the parameter space. The construction of an adapted local distance function is a topic itself and is addressed in Section 4. Another ingredient is the set of sample points \mathbb{S}_K whose construction is explained in the next section.

3.2. Construction of global set of sample points. This consists of the off-line stage of the algorithm. The first decision is to set the number N of basis functions we want to get involved in the on-line procedure (i.e. the size of the matrix system to be solved on-line).

A standard greedy algorithm, as described by Section 2, is then performed in a first stage until $N + 1$ basis functions are selected. The corresponding parameter values define the current set \mathbb{S}_{N+1} and the solutions $v(\boldsymbol{\mu})$, $\boldsymbol{\mu} \in \mathbb{S}_{N+1}$, are saved.

¹Note that N has not the same meaning as in the previous section any more; hence the cardinality of \mathbb{S}_K is not N but a $K > N$ generally much larger.

Then, the second stage of the algorithm is performed to further enrich \mathbb{S}_{N+1} . The construction of the further basis functions require the data of a distance function (as in the on-line stage). As for now we assume that the distance function is available, its recursive construction is explained in Section 4 (note that, by default, we use the Euclidean isotropic metric). We note that we can at each iteration and for each parameter value $\boldsymbol{\mu} \in \mathbb{P}$, compute an approximation with the N basis functions that lie in $B_{\boldsymbol{\mu}}$ and estimate the error by computing $\eta(\boldsymbol{\mu}; \mathbb{W}_{\boldsymbol{\mu}})$. The next parameter value(s) is chosen as the one corresponding to the worse approximation properties. It is important to notice that using local approximation spaces, several basis functions can be chosen per iteration. The following procedure is proposed.

At the end of a given iteration where we dispose of $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K$, a new iteration may start. The next sample point, $\boldsymbol{\mu}_{K+1}$, is chosen such that the estimated error $\eta(\boldsymbol{\mu}; \mathbb{W}_{\boldsymbol{\mu}})$ is maximized over the parameter space as in the traditional setting except that only a local approximation space $\mathbb{W}_{\boldsymbol{\mu}}$ is considered to compute the error (-estimator). Thus $\boldsymbol{\mu}_{K+1} = \arg \max_{\boldsymbol{\mu} \in \Xi_{\text{train}}} \eta(\boldsymbol{\mu}; \mathbb{W}_{\boldsymbol{\mu}})$. Next, the maximum estimated error $\eta(\boldsymbol{\mu}; \mathbb{W}_{\boldsymbol{\mu}})$ is searched over all parameter values lying outside of the domain of influence of $\boldsymbol{\mu}_{K+1}$. By the domain of influence of $\boldsymbol{\mu}_{K+1}$ we mean

$$\mathcal{D}_{\boldsymbol{\mu}_{K+1}} = \{\boldsymbol{\mu} \in \mathbb{P} \mid \boldsymbol{\mu}_{K+1} \in B_{\boldsymbol{\mu}}\} \cup B_{\boldsymbol{\mu}_{K+1}}.$$

Outside of the domain of influence of $\boldsymbol{\mu}_{K+1}$ more basis functions can however be added. This condition assures that at most one sample point is added per ball. This procedure is repeated, and more and more parts of the parameter domain are excluded, until the remaining set is the empty set. The practical implementation of this procedure is given the following pseudo-code.

EnrichmentLoop

Input: K, \mathbb{S}_K

Output: updated K, \mathbb{S}_K

1. Set $\widetilde{\Xi}_{\text{train}} = \Xi_{\text{train}}$, **err** = $\max_{\boldsymbol{\mu} \in \widetilde{\Xi}_{\text{train}}} \eta(\boldsymbol{\mu}; \mathbb{W}_{\boldsymbol{\mu}})$.
2. While $\widetilde{\Xi}_{\text{train}} \neq \emptyset$ and **err** > **tol**
3. Set $\boldsymbol{\mu}^{K+1} = \arg \max_{\boldsymbol{\mu} \in \widetilde{\Xi}_{\text{train}}} \eta(\boldsymbol{\mu}; \mathbb{W}_{\boldsymbol{\mu}})$, **err** = $\max_{\boldsymbol{\mu} \in \widetilde{\Xi}_{\text{train}}} \eta(\boldsymbol{\mu}; \mathbb{W}_{\boldsymbol{\mu}})$.
4. Compute $v(\boldsymbol{\mu}^{K+1})$ and set $\mathbb{S}_{K+1} = \mathbb{S}_K \cup \{\boldsymbol{\mu}^{K+1}\}$.
5. Set $\widetilde{\Xi}_{\text{train}} := \widetilde{\Xi}_{\text{train}} \setminus \{\boldsymbol{\mu} \in \Xi_{\text{train}} \mid \boldsymbol{\mu} \in B_{\boldsymbol{\mu}^{K+1}} \text{ or } \boldsymbol{\mu}^{K+1} \in B_{\boldsymbol{\mu}}\}$.
6. Set $K := K + 1$.
7. End while.

Algorithm 2: Basis enrichment loop.

4. ANISOTROPY AND DISTANCE FUNCTION - A GENERAL FINITE DIFFERENCE BASED APPROACH

In order to equip the parameter space \mathbb{P} with an appropriate metric to $\mathbb{W}_{\mathbb{P}}$ and distance function, one has to estimate, on the fly, the way the (unknown) function $v(\boldsymbol{\mu}) \in \mathbb{W}$ depends on $\boldsymbol{\mu}$. In this manner we aim to reduce at most the number of elements involved in the on-line reduced basis approximation.

Indeed, for instance assuming that the parameter domain \mathbb{P} is a subset of \mathbb{R}^2 , so that $\boldsymbol{\mu} = (\mu_1, \mu_2)$, and assuming that there exists a function φ of one variable, such that $v(\boldsymbol{\mu}) = \varphi(\mu_1 + \mu_2)$, then the optimal selection of parameters should be

sought e.g. along the line $\mu_1 = \mu_2$ considering only a one dimensional parameter. Of course the fact that v has such a behavior is generally not known. This has to be figured out from prior computations, during the greedy algorithm. This is actually a quite classical quest in numerical analysis, and this valuable knowledge leads to a large computational reduction. As in other instances, e.g. adaptive approaches like the one used in finite element approximations (see e.g. [3]), the construction of such a distance function $d(\cdot, \cdot)$ that accounts for the anisotropic behavior of changes of $v(\boldsymbol{\mu})$ with respect to variations in $\boldsymbol{\mu}$. Similarly as in other approaches, the distance function is derived from the knowledge of the Hessian of the function v with respect to the parameter $\boldsymbol{\mu}$.

We first propose a general framework to obtain an approximation of the Hessian by finite differences and then give the definition of the distance function. In what follows, we always assume that Ξ_{train} is such that its convex hull spans the parameter space \mathbb{P} .

4.1. Definition of the Hessian. The goal is to define a Hessian matrix $H(\boldsymbol{\mu})$ for each point $\boldsymbol{\mu} \in \Xi_{\text{train}}$ upon which the distance function will be based on. Since the Hessian is based on the reduced basis approximation, it is updated/constructed at each iteration in the off-line greedy algorithm (Section 3.2). To do so, compute a reduced basis approximation $P_{\boldsymbol{\mu}}(v(\boldsymbol{\mu}))$ of $v(\boldsymbol{\mu})$ based on the current local reduced basis, as explained in Section 3.1, on the stencil $\boldsymbol{\mu}(\alpha) = \boldsymbol{\mu} + \sum_{i=1}^p \alpha^i \delta \mu^i$, with $\boldsymbol{\mu} \in \Xi_{\text{train}}$ and where $\delta \mu^i$, $i = 1, \dots, p$, is a positive small increment in the i^{th} direction in the parameter and $\alpha^i = -1, 0, 1$, such that, at most, two of them are non zero. Let $\{\varphi_n\}_{n=1}^N$ denotes the basis of $\mathbb{W}_{\boldsymbol{\mu}}$ used to build $P_{\boldsymbol{\mu}}(v(\boldsymbol{\mu}))$ such that

$$P_{\boldsymbol{\mu}}(v(\boldsymbol{\mu})) = \sum_{n=1}^N v_n(\boldsymbol{\mu}) \varphi_n(\boldsymbol{x}).$$

From the 3^p approximations $P_{\boldsymbol{\mu}}(v(\boldsymbol{\mu}(\alpha)))$ of $v(\boldsymbol{\mu}(\alpha))$ on the surrounding stencil $\boldsymbol{\mu}(\alpha)$ of $\boldsymbol{\mu}$, one can, by finite differences, define a Hessian matrix with value in \mathbb{R} as follows

$$H_{ij}(\boldsymbol{\mu}) = \sum_{n=1}^N v_n(\boldsymbol{\mu}) D_{ij} v_n(\boldsymbol{\mu}),$$

where

$$D_{ij} v_n(\boldsymbol{\mu}) = \frac{v_n(\boldsymbol{\mu} + \delta \mu^i + \delta \mu^j) - v_n(\boldsymbol{\mu} - \delta \mu^i + \delta \mu^j) - v_n(\boldsymbol{\mu} + \delta \mu^i - \delta \mu^j) + v_n(\boldsymbol{\mu} - \delta \mu^i - \delta \mu^j)}{4 \delta \mu^i \delta \mu^j}.$$

The term $D_{ij} v_n(\boldsymbol{\mu})$ describes the anisotropy of the n -th mode while $v_n(\boldsymbol{\mu})$ indicates its weight in the approximation.

Remark 4.1. In the case one is only interested in a scalar output functional $s : \mathbb{W} \rightarrow \mathbb{R}$ of the solution, i.e. $s(v(\boldsymbol{\mu}))$ approximated by $s(v_{\text{rb}}(\boldsymbol{\mu}))$, one can also consider the Hessian do be defined as

$$H_{ij}(\boldsymbol{\mu}) = D_{ij} s(v_{\text{rb}}(\boldsymbol{\mu})).$$

Remark 4.2. The current construction of the Hessian based on a finite difference scheme involves quite a large number of points when the dimension p becomes too large. Alternative exist based only on the only points in \mathbb{S}_K that need still to be investigated and improved.

At this level we dispose for any $\boldsymbol{\mu} \in \mathbb{P}$ of a method to compute $H_{ij}(\boldsymbol{\mu})$. It remains to explain how we construct the metric tensor M from the knowledge of the Hessian matrix H . Indeed, given $\boldsymbol{\mu} \in \mathbb{P}$, we proceed in a standard manner and perform an eigenvalue decomposition

$$H(\boldsymbol{\mu}) = V \Lambda V^T$$

where V is an orthonormal matrix and $(\Lambda)_{ii} = \lambda_i$ a diagonal matrix consisting of the eigenvalues λ_i . Consider the diagonal matrix

$$|\Lambda|_{ii} = |\lambda_i|, \quad i = 1, \dots, p,$$

and the associated symmetric semi-positive definite matrix $M(\boldsymbol{\mu}) = V |\Lambda| V^T$ to define the metric tensor.

4.2. Construction of the distance function. The construction of the distance between two parameter points $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{P}$, denoted by $d(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$, in the metric space given by the metric tensor M is ideally defined by the length of the geodesic curve linking $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. Indeed, if $\boldsymbol{\gamma} : [0, 1] \rightarrow \mathbb{P}$ is a parametrization of the geodesic from $\boldsymbol{\mu}_1$ to $\boldsymbol{\mu}_2$ we would like to compute

$$\int_0^1 \sqrt{(\boldsymbol{\gamma}'(t))^T M(\boldsymbol{\gamma}(t)) \boldsymbol{\gamma}'(t)} dt.$$

The above integral is however not computable due to the lack of knowledge of the geodesic curve. The geodesic curve itself could be approximated, but this would slow down the on-line computation drastically. We therefore propose to approximate the above integral roughly by the following trapezoidal rule to define the distance

$$d(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \frac{1}{2} \sqrt{(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T M(\boldsymbol{\mu}_1) (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)} + \frac{1}{2} \sqrt{(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T M(\boldsymbol{\mu}_2) (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)}$$

where we also took the approximated $\boldsymbol{\gamma}'(t) \approx (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$ for $t = 0, 1$ (which is correct when dealing with the Euclidian metric).

We note that the construction of the Hessian $H(\boldsymbol{\mu})$ and thus the distance function depends on the approximation $P_{\boldsymbol{\mu}}(v(\boldsymbol{\mu}))$, and thus on \mathbb{S}_K . That is, for each reduced basis model we can define a distance function and we recursively update the distance function at each step of the stage 2 of the greedy algorithm. Therefore, the distance function is, simultaneously with $P_{\boldsymbol{\mu}}(v(\boldsymbol{\mu}))$, converging to a final distance function that is used in the on-line procedure.

Let us make the following comments:

- [1] It is easy to see that our distance function is symmetric by construction.
- [2] The matrix $M(\boldsymbol{\mu})$ is semi-definite positive. Allowing for zero eigenvalues is indeed a desired property since it allows to shrink certain directions in parameter space where there is no variation in the solution with respect to the parameter.
- [3] Defining a mesh consisting of p -dimensional simplices with vertices being the set Ξ_{train} allows to interpolate/represent quantities only defined on Ξ_{train} by means of piecewise linear functions on the mesh for any parameter value $\boldsymbol{\mu} \in \mathbb{P}$, since the convex hull of Ξ_{train} coincides with the parameter space \mathbb{P} . We can therefore easily access the interpolated metric tensor $M(\boldsymbol{\mu})$, for any $\boldsymbol{\mu} \in \mathbb{P}$, by means of its interpolated component functions $M_{ij}(\boldsymbol{\mu})$.

Remark 4.3. *As an alternative to interpolating the coefficients in a p -dimensional space one can construct the Hessian on-line as explain above for any $\mu \in \mathbb{P}$.*

5. ADAPTIVE TRAINING SETS BASED ON HESSIAN

The goal of this section is to present an approach that further allows to reduce the off-line computational cost. Here, we focus on the training set Ξ_{train} . It is aimed to keep its cardinality as small as possible, but large enough to capture the local geometry of the parametrized system. We aim to give an answer to the following question: Given the computational resources of handling Q_M points in Ξ_{train} , how do we place them optimally in the parameter space?

We propose to place the Q_M points uniformly in \mathbb{P} with respect to the empirically constructed metric tensor $M(\mu)/r(\mu)^2$. The factor $1/r(\mu)^2$ is motivated by the fact that a ball defined in (1) corresponds (approximatively) to a unit ball in the metric defined by $M(\mu)/r(\mu)^2$. Since the metric gets updated at each iteration of the greedy algorithm, we also need to adapt the training points Ξ_{train} at each iteration. Further, we propose that the cardinality of Ξ_{train} is an increasing function of the inverse of the actual error (-estimation) **err**. Let Q_m be the minimal cardinality of Ξ_{train} at the beginning of the greedy algorithm and Q_M the maximal target cardinality at the end of the algorithm once the tolerance criteria **err** < **tol** is reached, then we define

$$Q(\text{err}) = \left\lceil \frac{Q_M - Q_m}{\log(\text{tol})} \log(\text{err}) + Q_m \right\rceil$$

to be the number of points at the next iteration in Ξ_{train} .

To summarize, the framework that we propose is that the training set Ξ_{train} of cardinality $Q(\text{err})$ is constructed such that the points are uniformly distributed in the empirically constructed metric tensor $M(\mu)/r(\mu)^2$.

Our implementation of this concept is based on the Delaunay mesh defined by the points of Ξ_{train} . We use the software FreeFem++ [9] which provides a uniform mesh with respect to $M(\mu)$. Other implementations are of course also possible.

Remark 5.1. *Enriching the training set during the sampling process is not a new idea. An adaptively enriching greedy sampling strategy was proposed in [7] however it is not based, as it is the case in the present paper, on a learning process of the geometry of the manifold. In addition our approach is more precise since it not only detects where the manifold of the solutions (as a function of the parameters) is complex but also the directions that are important to sample and those that are not. Second in our approach the size of the discrete system to be solved on-line is determined by the user leading to a controlled simulation cost.*

6. ON-LINE ORTHONORMALIZATION OF BASIS FUNCTIONS

In the standard approach, it is well known that the stable implementation of the reduced basis method requires the orthonormalization of the snapshots $v(\mu)$, $\mu \in \mathbb{S}_N$ [13]. This orthonormalization is prepared off-line and the vectorial space \mathbb{W}_N spanned by the series $\{v(\mu) \mid \mu \in \mathbb{S}_N\}$ is not affected by this change of basis.

In our alternative method, there are a large number of approximation spaces, almost one for each μ : $\mathbb{W}_\mu = \text{span}\{v(\tilde{\mu}) \mid \tilde{\mu} \in \mathbb{S}_\mu\}$. We cannot compute all various orthonormalized basis sets for all possible \mathbb{W}_μ off-line. This would be too costly and using way too much memory.

Nevertheless, we can still *prepare* the orthonormalization process so that, the on-line orthonormalization is feasible. For any $\hat{\boldsymbol{\mu}}_i$ and $\hat{\boldsymbol{\mu}}_j$ in \mathbb{S}_K , let us compute off-line the scalar products

$$(2) \quad M_{ij} = \langle \hat{v}(\boldsymbol{\mu}_j), v(\hat{\boldsymbol{\mu}}_i) \rangle, \quad i, j = 1, \dots, K$$

where $\langle \cdot, \cdot \rangle$ represents an appropriate scalar product : say either L^2 or H^1 type. Note that we can even compute only those scalar products corresponding to pairs of parameters $(\hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\mu}}_j)$ that are close, indeed if $\hat{\boldsymbol{\mu}}_i$ and $\hat{\boldsymbol{\mu}}_j$ are distant in \mathbb{P} , they will never be in the same $\mathbb{S}_{\boldsymbol{\mu}}$ for any $\boldsymbol{\mu} \in \mathbb{P}$.

In the on-line stage, once the approximation in $\mathbb{W}_{\boldsymbol{\mu}}$ will be required, we can orthonormalize the basis set $\{v(\tilde{\boldsymbol{\mu}}) \mid \tilde{\boldsymbol{\mu}} \in \mathbb{S}_{\boldsymbol{\mu}}\}$ first by declaring an order in the parameters

$$(3) \quad \{\tilde{\boldsymbol{\mu}} \in \mathbb{S}_{\boldsymbol{\mu}}\} = \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_N\}$$

then perform a classical Gram-Schmidt orthonormalization process

$$(4) \quad \begin{aligned} \zeta_1 &= \beta_{11}v(\boldsymbol{\mu}_1), \\ \zeta_2 &= \beta_{22}v(\boldsymbol{\mu}_2) + \beta_{21}\zeta_1, \\ &\vdots \\ \zeta_N &= \beta_{NN}v(\boldsymbol{\mu}_N) + \sum_{i=1}^{N-1} \beta_{Ni}\zeta_i, \end{aligned}$$

where the coefficients β_{ni} are chosen so that ζ_n is orthogonal to

$$\text{span}\{v(\boldsymbol{\mu}_i) \mid i = 1, \dots, n-1\} = \text{span}\{\zeta_i \mid i = 1, \dots, n-1\}$$

and the coefficients β_{nn} are chosen so that ζ_n is norm 1. It is well known that these coefficients exist, are unique and that, in order to compute them, we have to know each scalar product $\langle v(\boldsymbol{\mu}_n), \zeta_i \rangle$, $i = 1, \dots, n-1$ that can be easily computed mimicking the Gram-Schmidt procedure (4) but using the precomputed coefficients M_{ij} . For the construction of the coefficients β_{nj} , we proceed as follows (it helps to have in mind that $\tilde{\beta}_{kn}$ stands for $-\langle v(\boldsymbol{\mu}_k), \zeta_n \rangle$) : for all $n = 1, \dots, N$ apply recursively

$$(5) \quad \begin{aligned} \alpha &= \left(M_{nn} - \sum_{j=1}^{n-1} |\tilde{\beta}_{nj}|^2 \right)^{\frac{1}{2}}, \\ \beta_{nn} &= \frac{1}{\alpha}, \\ \beta_{nj} &= \frac{\tilde{\beta}_{nj}}{\alpha}, \quad k = 1, \dots, n-1, \\ \tilde{\beta}_{kn} &= -\beta_{nn} M_{nk} + \sum_{j=1}^{n-1} \beta_{nj} \tilde{\beta}_{kj}, \quad k = 1, \dots, N. \end{aligned}$$

Note that the construction of α is not subject to instabilities since the sum involves only positive values.

Next, we notice that the basis functions $\{\zeta_n\}_{n=1}^N$ can alternatively be expressed by the following linear transformation

$$\begin{aligned} \zeta_1 &= \gamma_{11}v(\boldsymbol{\mu}_1), \\ \zeta_2 &= \gamma_{22}v(\boldsymbol{\mu}_2) + \gamma_{21}v(\boldsymbol{\mu}_1), \end{aligned}$$

$$(6) \quad \begin{array}{c} \vdots \\ \zeta_N \end{array} = \sum_{i=1}^N \gamma_{Ni} v(\boldsymbol{\mu}_i),$$

from the set of basis functions $\{v(\boldsymbol{\mu}_n)\}_{n=1}^N$. It is our final intention to construct the coefficients γ_{ni} representing the change of basis functions from $\{v(\boldsymbol{\mu}_n)\}_{n=1}^N$ to $\{\zeta_n\}_{n=1}^N$ as defined by (6) in a stable way and without any operation depending on the length of the vectors representing $v(\boldsymbol{\mu}_n)$ to guarantee stability and good on-line performance.

Remark 6.1. *Note that the coefficients $\gamma_{n,i}$ could be obtained based on (6) directly. Indeed, each ζ_n is orthogonal to $\text{span}\{v(\boldsymbol{\mu}_i) \mid i = 1, \dots, n-1\}$ allowing to compute the coefficients γ_{ni} , $i = 1, \dots, n-1$. This however involves solving a linear systems based on the mass matrix with coefficients $M_{ij} = \langle v(\boldsymbol{\mu}_j), v(\boldsymbol{\mu}_i) \rangle$, $i, j = 1, \dots, n-1$, which can be ill conditioned due to the fact that the set of functions $\{v(\boldsymbol{\mu}_i)\}_i$ is heavily linearly dependent.*

The set of coefficients $\{\gamma_{ni}\}_{n,i=1}^N$ can be obtained by $\mathcal{O}(N^3)$ operations from the coefficients β_{nj} , through a triangular process. Indeed, the coefficients γ_{ni} can be obtained by the following recursive formula

$$(7) \quad \gamma_{ni} = \begin{cases} \sum_{k=1}^{n-1} \beta_{nk} \gamma_{ki} & \text{if } i < n, \\ \beta_{nn} & \text{if } i = n, \\ 0 & \text{if } i > n, \end{cases} \quad \text{for all } n = 1, \dots, N.$$

As mentioned above, the set of basis functions $\{\zeta_n\}_{n=1}^N$ leads to well-conditioned matrices, the precomputed quantities are nevertheless expressed in the basis $\{v(\boldsymbol{\mu}_n)\}_{n=1}^N$, but the change of basis functions is described by the coefficients $\{\gamma_{ni}\}_{n,i=1}^N$ as stated in equation (6).

As an illustration assume that the reduced basis problem to be solved is about the approximation of the solution to a parametrized PDE, written in a variational form as : for a given parameter value $\boldsymbol{\mu}$, find $u(\boldsymbol{\mu}) \in \mathbb{W}_{\boldsymbol{\mu}}$ such that

$$a(u(\boldsymbol{\mu}), v; \boldsymbol{\mu}) = f(v; \boldsymbol{\mu}), \quad \forall v \in \mathbb{W}_{\boldsymbol{\mu}},$$

where it is assumed, for the sake of simplicity that a and f are affine decomposable

$$(8) \quad a(w, v; \boldsymbol{\mu}) = \sum_{q=1}^{Q_a} g_q(\boldsymbol{\mu}) a_q(w, v), \quad f(v; \boldsymbol{\mu}) = \sum_{q=1}^{Q_f} h_q(\boldsymbol{\mu}) f_q(v).$$

The ultimate goal is to construct the well-conditioned stiffness matrix $a(\zeta_j, \zeta_i; \boldsymbol{\mu})$ derived from the off-line pre-computation of the series $A_{ij}^q = a_q(v(\boldsymbol{\mu}_j), v(\boldsymbol{\mu}_i))$, for $1 \leq i, j \leq N$. To do so, follow the next steps:

- [1] Compute the coefficients $\{\beta_{ni}\}_{n,i=1}^N$ following (5).
- [2] Compute the coefficients $\{\gamma_{ni}\}_{n,i=1}^N$ following (7).
- [3] For each $q = 1, \dots, Q_a$, the matrix $a_q(\zeta_j, \zeta_i)$ can be computed by $a_q(\zeta_j, \zeta_i) = \sum_{k,l=1}^N \gamma_{jk} \gamma_{il} A_{lk}^q$.
- [4] Build the solution matrix by summing $a(\zeta_j, \zeta_i; \boldsymbol{\mu}) = \sum_{q=1}^{Q_a} g_q(\boldsymbol{\mu}) a_q(\zeta_j, \zeta_i)$.

Similarly, the evaluation of each component of the vector $f(\zeta_i; \boldsymbol{\mu})$, $1 \leq i \leq N$ can be performed in $\mathcal{O}(Q_f N)$ operations, and the inversion of the associated matricial problem is done in $\mathcal{O}(N^3)$ further operations.

7. OVERVIEW OF THE FRAMEWORK

Before we start with the numerical tests, an abstract description of both the off-line and on-line parts of the locally adaptive greedy algorithm is given in the following box:

LocallyAdaptiveGreedy Off-line Input: N Output: $K, \mathbb{W}_K, \mathbb{S}_K$
<pre> ===== Stage 1 ===== 1. Perform a classical greedy algorithm to select $N + 1$ basis functions. ===== Stage 2 ===== 2. For each $\boldsymbol{\mu} \in \Xi_{\text{train}}$, do (parameter space search) a) Compute the error estimate $\eta(\boldsymbol{\mu}, \mathbb{W}_{\boldsymbol{\mu}})$. b) Update the Hessian matrix $H_{ij}(\boldsymbol{\mu})$. 4. Enrich the set of basis functions (EnrichmentLoop, Section 3.2). 5. Create a new training set Ξ_{train} (Section 5). 6. Go to 2. until tolerance tol is achieved. Algorithm 3: Locally adaptive greedy algorithm (Off-line).</pre>

LocallyAdaptiveGreedy On-line Input: $K, \mathbb{W}_K, \mathbb{S}_K, \boldsymbol{\mu}$ Output: $P_{\boldsymbol{\mu}}(v(\boldsymbol{\mu}))$
<pre> 1. Find the approximation space $\mathbb{W}_{\boldsymbol{\mu}}$: a) Compute the distance $d(\boldsymbol{\mu}, \tilde{\boldsymbol{\mu}})$ for each $\tilde{\boldsymbol{\mu}} \in \mathbb{S}_K$ (Section 4.1). b) Choose the N closest $\tilde{\boldsymbol{\mu}} \in \mathbb{S}_K$ to form $\mathbb{S}_{\boldsymbol{\mu}}$ and $\mathbb{W}_{\boldsymbol{\mu}} = \text{span}\{v(\tilde{\boldsymbol{\mu}}) \mid \tilde{\boldsymbol{\mu}} \in \mathbb{S}_{\boldsymbol{\mu}}\}$ (Section 3.1). 2. Orthonormalize the basis functions $\{v(\tilde{\boldsymbol{\mu}}) \mid \tilde{\boldsymbol{\mu}} \in \mathbb{S}_{\boldsymbol{\mu}}\}$ to build a new basis $\{\xi_1, \dots, \xi_N\}$ (Section 6). 3. Compute the projection $P_{\boldsymbol{\mu}}(v(\boldsymbol{\mu}))$ based on the basis $\{\xi_1, \dots, \xi_N\}$. Algorithm 4: Locally adaptive greedy algorithm (On-line).</pre>

8. NUMERICAL RESULTS USING THE L^2 -PROJECTION

Before we pass to a full Reduced Basis model based on a parametrized Galerkin-projection, we use a parametrized family of explicitly given functions combined with the L^2 -projection to define the projection operator P_N in order to test our framework. Reminding Cea's lemma which is a consequence of the Galerkin approach, studying the best-approximation is sufficient to test the method of locally adaptive approximation spaces. The error is computed in an exact manner and no *a posteriori* estimation is used.

Further, in a first step, Section 8.1, we only test the feature of local anisotropic approximation balls without adapting the training set using the distance function.

That is we use the same fixed Ξ_{train} throughout all steps of the algorithm. In a second step, Section 8.2, we test the entire algorithm.

For all numerical tests, Ω and \mathbb{P} are discretized by a regular lattice of 75×75 points and the target tolerance is set to 10^{-4} if not otherwise mentioned.

8.1. Numerical results with fixed training set.

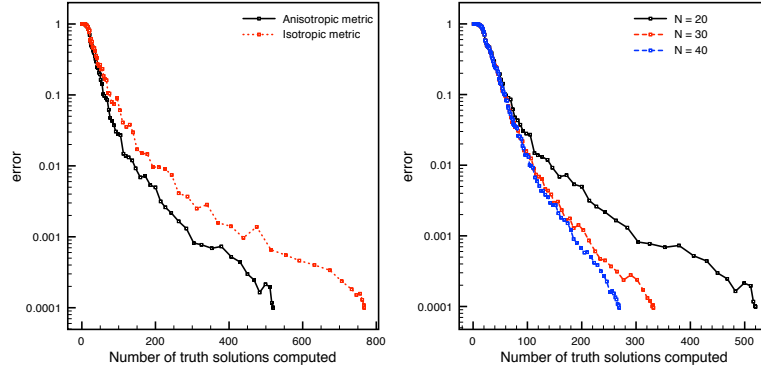
Test 1. We start with presenting a numerical example to illustrate the benefit of the local anisotropic approximation spaces. Consider the function

$$f_1(\mathbf{x}; \boldsymbol{\mu}) = \exp \left[\frac{-(x_1 - 0.1(\mu_1 - \mu_2))^2}{0.01} - \frac{(x_2 - (\mu_1 + \mu_2))^2}{0.01} \right],$$

$$\mathbf{x} \in \Omega = (-1, 1)^2, \quad \boldsymbol{\mu} \in \mathbb{P} = [-0.5, 0.5]^2$$

that exhibits a constant anisotropy of parameters over the whole parameter space. The dependency in the $(\mu_1 + \mu_2)$ -direction is ten times stronger than in the $(\mu_1 - \mu_2)$ -direction. Figure 1 (left) illustrates the evolution of the number of basis functions that are necessary to be computed versus the achieved accuracy in the L^∞ -norm for $N = 20$ for the anisotropic approach compared to the same algorithm but using isotropic local approximation spaces as comparison. One can observe that using the anisotropic approach is beneficial in terms of the number of truth solutions to be computed at the off-line stage. Figure 1 (right) presents the same quantity but for varying $N = 20, 30, 40$. Not surprisingly the number of needed basis functions decreases while increasing N and more and more the exponential convergence establishes as is manifested for the standard greedy algorithm in this example. We can also observe that for a lower N more basis functions can be included per iteration.

Figure 2 illustrates the local approximation spaces expressed in \mathbb{P} , the radius as a function of the parameter and the sample set \mathbb{S}_K for the anisotropic and the isotropic approach.



Comparison of different approaches.

Comparison under variation of N .

FIGURE 1. Test 1: Accuracy with respect to the number of truth solutions to be computed in comparison with the isotropic approach for $N = 20$ (left) and the for different values of $N = 20, 30, 40$ (right).

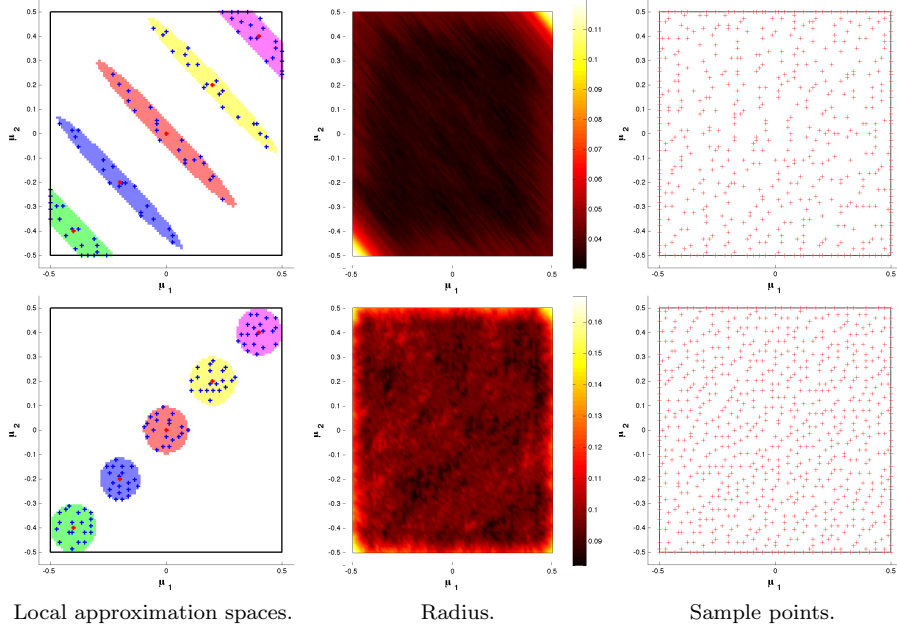


FIGURE 2. Test 1: Local approximation spaces for selected parameter values (left), radius as a function of the parameters (middle) and sample points (right) for anisotropic approach (top) and the isotropic version (bottom, as comparison) with $N = 20$.

Test 2. The previous example is in some sense idealized since the anisotropy is not varying throughout the parameter domain. In this regard, the next example

$$f_2(\mathbf{x}; \boldsymbol{\mu}) = \exp \left[-\frac{(x_1 - (\mu_1^2 + \mu_2^2))^2}{0.01} - \frac{(x_2 - (\mu_1^2 + \mu_2^2))^2}{0.01} \right],$$

$$\mathbf{x} \in \Omega = (-1, 1)^2, \quad \boldsymbol{\mu} \in \mathbb{P} = [-0.5, 0.5]^2$$

is more interesting. It presents a family of parametrized functions where the functions (as functions of \mathbf{x}) are constant along concentric circles around the origin in parameter space. As an example, all four corners of the parameter space define the same function.

We compare again the number of required truth solutions for the anisotropic approach compared with the presented algorithm but using isotropic approximation spaces, which is presented in Figure 3 (left) for $N = 5$. We observe that 60% of the number of truth solutions to be computed can be saved by constructing the problem-dependent approximation spaces. The corresponding local approximation spaces, radii and sample points \mathbb{S} are illustrated in Figure 4. Figure 3 (right) also presents the number of required truth solutions for $N = 5$ but for different tolerance levels. We observe that starting from a critical value N_c of computed truth solutions, that depends on the tolerance, the curves show different behavior. During the iterations before N_c the complete parameter region has to be scanned because the tolerance is achieved nowhere in the parameter space. During the iterations after N_c however the tolerance is achieved in an increasing region of the parameter space and this

part needs no longer to be not scanned. The enrichment of new truth solutions is then limited in space and the final phase of the greedy algorithm converges faster.

Additionally we present in Figure 5 the local approximation spaces for the different tolerance levels. One can clearly see how the approximation spaces take into account the local geometry that also allow non-convex approximation spaces.

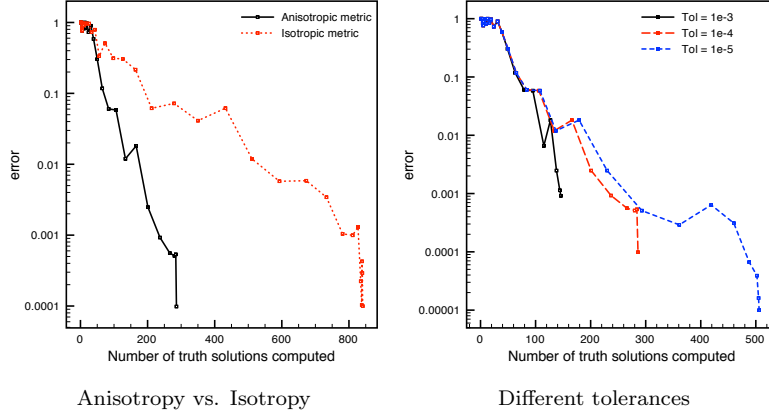


FIGURE 3. Test 2: Accuracy with respect to the number of truth solutions to be computed in comparison with the isotropic approach (left) and the for different end tolerances (right). In both cases there holds $N = 5$.

Test 3. The next example

$$f_3(\mathbf{x}; \boldsymbol{\mu}) = \exp \left[-\frac{(x_1 - (\mu_1 + 3\mu_2))^2}{0.1 + 5|\mu_1 + 3\mu_2|} - \frac{(x_2 - (3\mu_1 - \mu_2))^2}{0.1 + 5|3\mu_1 - \mu_2|} \right],$$

$$\mathbf{x} \in \Omega = (-1, 1)^2, \quad \boldsymbol{\mu} \in \mathbb{P} = [-0.5, 0.5]^2$$

is interesting in the sense that it presents an almost singularity in parameter space at the origin. The performance of the anisotropic approach, compared to the isotropic version, is presented in Figure 6. Again, the anisotropic approaches outperform the isotropic one and about a 56% of computations of truth solutions can be saved.

The local approximation spaces, the radius and the sample points with $N = 10$ are presented in Figure 7. We observe *a posteriori* that the training space Ξ_{train} was not sampled fine enough. Indeed in the region around the origin (and the cross for the isotropic version), every training point is included in the set of sample points. This also explains the sudden drop of the convergence in Figure 6, in particular for the isotropic version. This is a known difficulty in greedy methods and our approach with adaptive training sets presents a solution to this problem. The corresponding numerical results are presented in the next section.

Further, we plot the local approximation spaces for different values of the tolerance levels in Figure 8. We observe that the originally non-connected approximation spaces are becoming connected with increased tolerance.

We also recognize that the scheme detects the cross where the behavior of the function is most singular and present in Figure 9 the sample points that were chosen

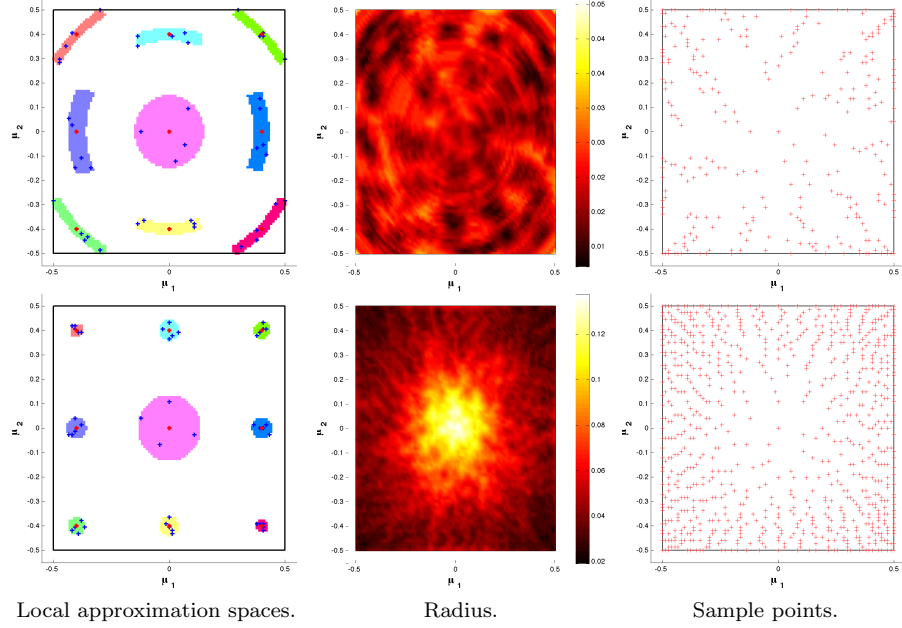


FIGURE 4. Test 2: Local approximation spaces for selected parameter values (left), radius as a function of the parameters (middle) and sample points (right) for the presented approach (top) in comparison with the isotropic version (bottom) for $N = 5$.

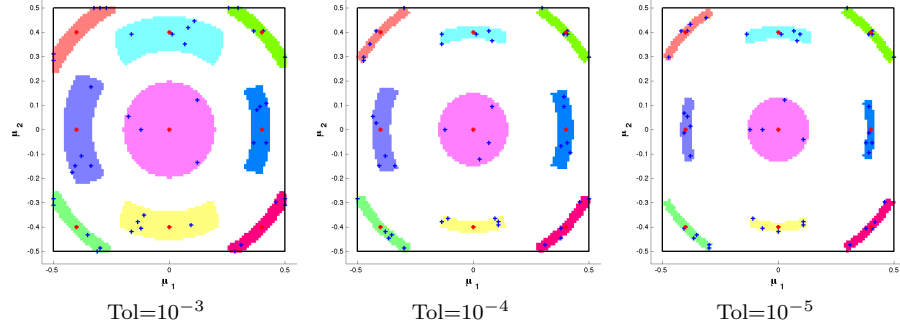


FIGURE 5. Test 2: Local approximation spaces for different tolerance levels.

for different functions of the type

$$f_{3,\xi_1,\xi_2}(\mathbf{x};\boldsymbol{\mu}) = \exp \left[-\frac{(x_1 - \xi_1(\boldsymbol{\mu}))^2}{0.1 + 5|\xi_1|} - \frac{(x_2 - \xi_2(\boldsymbol{\mu}))^2}{0.1 + 5|\xi_2|} \right]$$

$$\mathbf{x} \in \Omega = (-1, 1)^2, \quad \boldsymbol{\mu} \in \mathbb{P} = [-0.5, 0.5]^2$$

where ξ_1 and ξ_2 are functions of $\boldsymbol{\mu} = (\mu_1, \mu_2)$ using an error tolerance of 10^{-3} in this case.

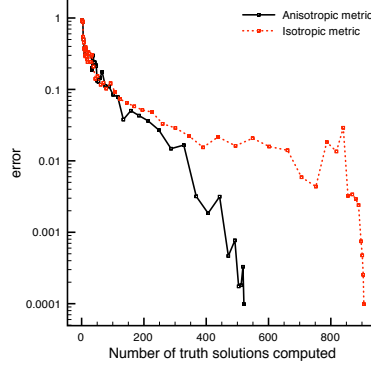


FIGURE 6. Test 3: Accuracy with respect to the number of truth solutions to be computed in comparison with the isotropic approach for $N = 10$.

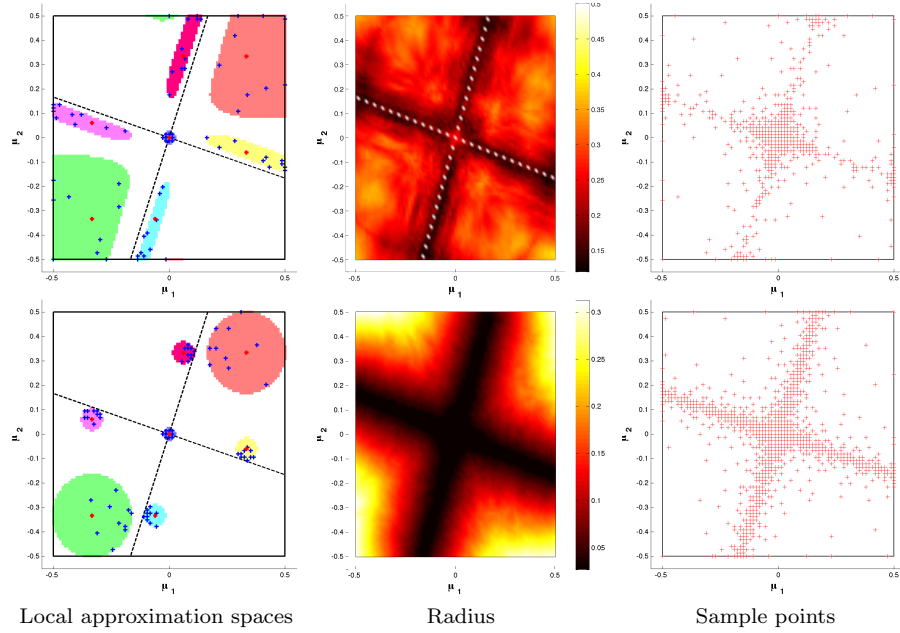


FIGURE 7. Test 3: Local approximation spaces for selected parameter values (left), radius as a function of the parameters (middle) and sample points (right) for the presented approach (top) in comparison with the isotropic version (bottom) for $N = 10$.

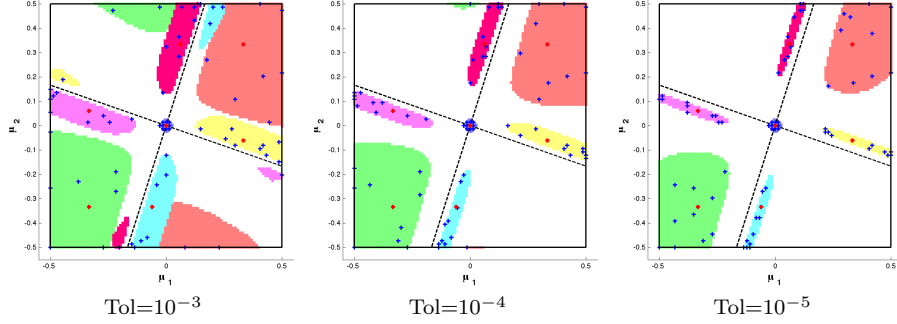


FIGURE 8. Test 3: Local approximation spaces for different tolerance levels.

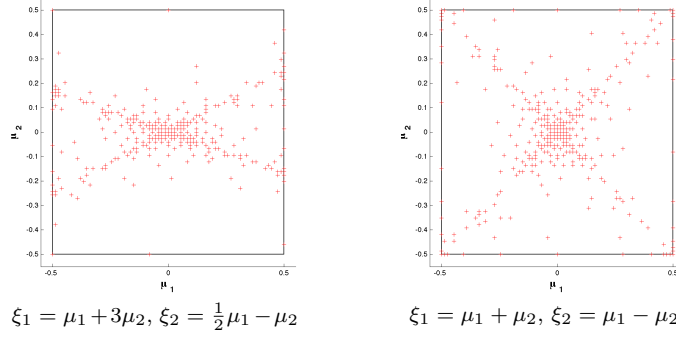


FIGURE 9. Sample set patterns for different type of parametrized functions with singular behavior on two crossing line with different slopes.

8.2. Numerical results with adapted training sets. In order to compare the accuracy of this version with adapted training sets we introduce a fixed test set of sample points Ξ_{test} . In the following examples it consists of a lattice of 75×75 points that corresponds to the fixed training set of the tests of the previous section. We use the command `adaptmesh` in FreeFem [9] in order to adapt the sample points by constructing a desired uniform mesh in the new metric defined by the Hessian.

The target advantage of our new approach with an adapting training set is not to use fewer computed solutions than the approach using a fixed training set, at least in low parameter's dimension but rather provide an improved reliability. Indeed a non-adapted training set may leave large errors in parameter's regions that are not sufficiently sampled by the fixed training set. On the contrary the adapted set tracks these unexplored parameter's regions where the RB approximation error remains large. The points of the training set are chosen wisely using the information available from the local geometry of the system. Note however that for problem where the manifold $\mathbb{W}_{\mathbb{P}}$ happens to be uniformly regular, the benefit of our approach will be on the computational efficiency as a result of the use of a small training set in the beginning that is gradually increasing.

Figure 10 illustrates the total number of basis functions that are necessary to be computed versus the achieved accuracy of the proposed algorithm with adaptive training sets for the test cases 1, 2 and 3. In each plot, we present the error of the version using an adaptive training set on the adaptive training set and the error evaluated on the fixed testing set Ξ_{test} . The accuracy of the version with a fixed training set (which equals Ξ_{test}) is given as comparison.

Test 1. From Figure 10 (left), one can observe that the error on the training set and on the test set of the approach with an adaptive training set are similar. This illustrates that the accuracy is not only guaranteed on the adaptive training set but also satisfied on the test grid. Both errors are also similar to the error of the algorithm using a fixed training set as presented in Section 8.1.

The adaptive version is as good as the version on a fixed training set, but uses less evaluation of the error estimate as is shown in the following table:

Number of training point evaluations using adaptive training sets:	150'923
Number of training point evaluations using fixed training sets:	455'625

We observe a gain of 65% fewer error evaluations.

Figure 11 presents the training set, local approximation spaces, radii and sample points for this calculation.

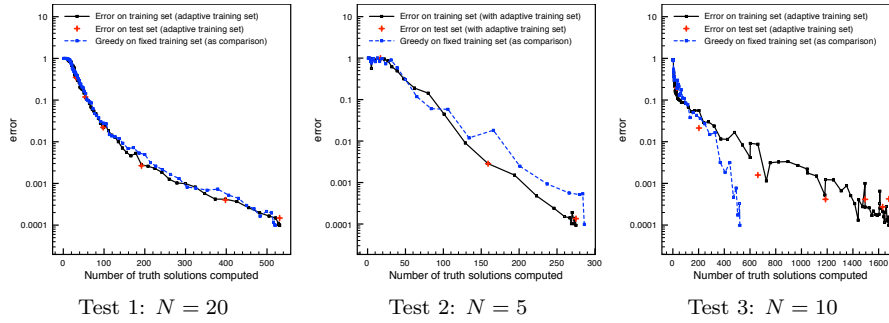


FIGURE 10. Accuracy with respect to the number of truth solutions that need to be computed for all three test case.

Test 2. Next, we consider the second example of the previous section. Figure 10 (middle) compares this version (with adaptive training sets) with the version of the previous section using a fixed training set. One can observe again that the performance is similar than the version using a fixed training set, and that the accuracy is also satisfied on the fixed test set Ξ_{test} .

Comparing again with the version using a fixed training set, the version with an adaptive training set requires less evaluation of the error estimate as is shown in the following table:

Number of training point evaluations using adaptive training sets:	60'202
Number of training point evaluations using fixed training sets:	151'875

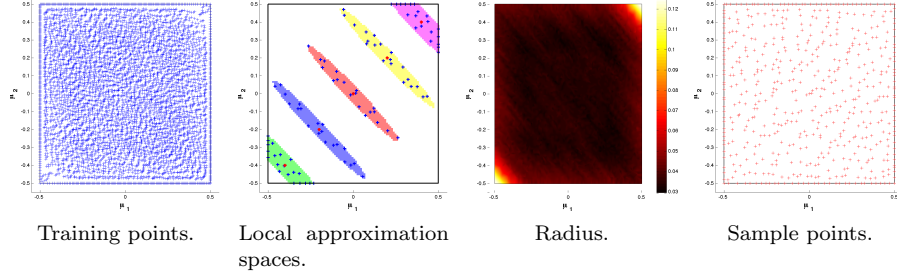


FIGURE 11. Test 1: Training points, local approximation spaces for some specific parameter values, radii and sample points for greedy algorithm with adaptive training set and $N = 20$.

We observe a gain of 70% fewer error evaluations.

Figure 12 presents the training set, local approximation spaces, radii and sample points of this calculation.

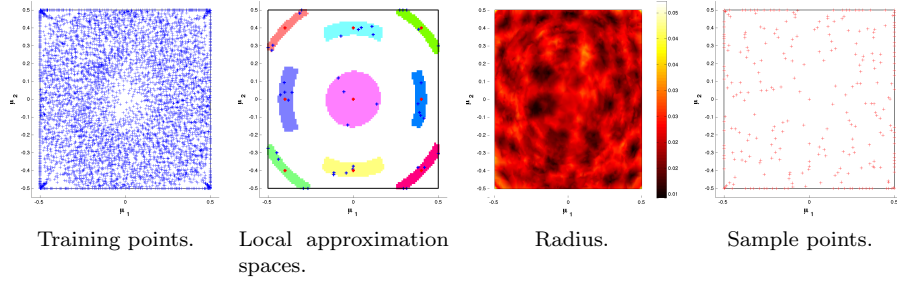


FIGURE 12. Test 2: Training points, local approximation spaces for some specific parameter values, radii and sample points for greedy algorithm with adaptive training set and $N = 5$.

Test 3. Finally, we apply the new algorithm with an adaptive training set to the third example. As already mentioned, the solution of this example shows very strong changes with respect to the parameter and the version with a fixed training set does not resolve this since the training set was not sampled fine enough. We expect the version with the adaptive training set to perform accurately on the entire parameter space, and thus in contrast using more sample points than the version with a fixed training set. Figure 10 (right) presents the evolution of the number of sample points versus the accuracy of the algorithm. One can observe that this approach (with adaptive training set) uses indeed more sample points than the version on a fixed training set as explained above. Finally, Figure 13 presents the training set, local approximation spaces, radii and sample points. We observe that the training points are in accordance with the sample points, and that a more dense sampling is indeed required around the origin. We observe that the local approximation spaces are now connected also for the tolerance of 10^{-4} .

In order to illustrate the benefit of using adaptive training sets also in this case we consider a test sample of 100×100 uniformly distributed points in the region $[-0.05, 0.05]^2$ around the origin. Figure 14 illustrates the error distribution using the online procedure generated using a fixed (left) and an adaptive (right) training set. The maximum error is 0.046 resp. $1.29 \cdot 10^{-4}$. While the error tolerance is almost satisfied in the latter case, it is clearly not the case for the former approach. Further, the number of error evaluations is given in the following table:

Number of training point evaluations using adaptive training sets:	181'672
Number of training point evaluations using fixed training sets:	247'500

Thus, the adaptive version uses still less error estimator evaluations and is more accurate in the region around the origin. Also, the error is more equally distributed.

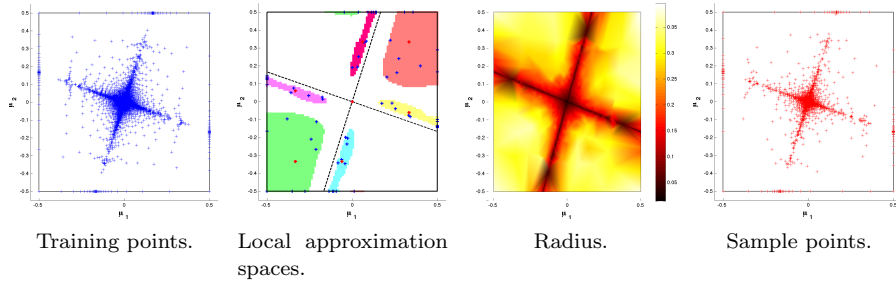


FIGURE 13. Test 3: Training points, local approximation spaces for some specific parameter values, radii and sample points for greedy algorithm with adaptive training set and $N = 10$.

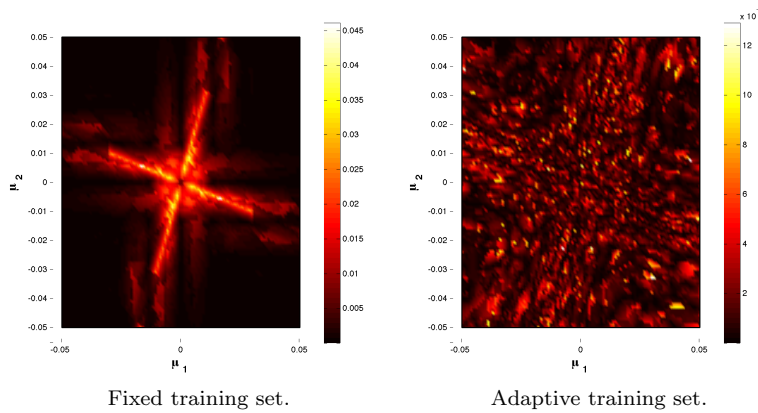


FIGURE 14. Test 3: Error distribution on a region around the origin of the greedy version using a fixed (left) and an adaptive (right) training set.

9. NUMERICAL RESULTS USING A GALERKIN-PROJECTION

Test 4. Finally, we present an example involving the full reduced basis framework. We consider a two-dimensional steady convection-diffusion equation on $\Omega = (0, 1)^2$. Homogeneous Dirichlet boundary conditions are imposed on the left, upper and right boundary, i.e. on $\Gamma_0 = \{(x, y) \in \partial\Omega \mid x = 0, \text{ or } x = 1, \text{ or } y = 1\}$. On the lower boundary, i.e. on $\Gamma_g = \{(x, 0) \in \partial\Omega \mid x \in [0, 1]\}$, we impose the following Dirichlet data

$$u(x, 0) = g(x) := x \chi_{[0, 0.5)} + (-40x + 21) \chi_{[0.5, 0.525]} + (x - 1) \chi_{[0.525, 1]},$$

(essentially a slightly regularized saw tooth function) where χ_X denotes the characteristic function for the mono-dimensional region X . Thus, the boundary condition presents a strong gradient between $(0.5, 0)$ and $(0.525, 0)$. Next, we introduce two parameters, the diffusion coefficient $\varepsilon = 10^{\mu_1}$ and the direction of the convection parametrized by an angle μ_2 :

$$\beta(\mu_2) = (\sin(\mu_2), \cos(\mu_2))^T.$$

Denoting the parameter vector by $\mu = (\mu_1, \mu_2) \in \mathbb{P} = [-2.5, 0] \times [-\pi/4, \pi/4]$, the parametrized problem writes as: for any $\mu \in \mathbb{P}$, find $u(\mu) \in H^1(\Omega)$ such that

$$(9) \quad -10^{\mu_1} \Delta u(\mu) + \nabla \cdot (\beta(\mu_2) u(\mu)) = 0, \quad \text{in } \Omega,$$

$$(10) \quad u(\mu) = g, \quad \text{on } \Gamma_g,$$

$$(11) \quad u(\mu) = 0, \quad \text{on } \Gamma_0.$$

The solutions presents a convected near discontinuity, that is also subject to a diffusion process, along a parametrized velocity field and parametrized diffusion coefficient. To solve the truth problem, we use a finite element solver using a uniform quadrilateral mesh and tensorized polynomials of degree 3 with mesh size $h = 0.025$. Different solutions corresponding to different parameter values are illustrated in Figure (15) and we highlight that the discretization is fine enough in order to avoid strong oscillations due to under-resolution related to the dominant convection for all parameter values. For the error computation/estimation that is used in the greedy algorithm, we use for sake of simplicity the exact error between the truth solution and the reduced basis approximation.

This problem can be cast as a variational problem with bilinear- and linear forms that are affine decomposable as described in (8) with

$$\begin{aligned} a_1(w, v) &= \int_{\Omega} \nabla w \nabla v \, d\mathbf{x}, & g_1(\mu) &= 10^{\mu_1}, \\ a_2(w, v) &= \int_{\Omega} \partial_x w v \, d\mathbf{x}, & g_2(\mu) &= \sin(\mu_2), \\ a_3(w, v) &= \int_{\Omega} \partial_y w v \, d\mathbf{x}, & g_3(\mu) &= \cos(\mu_2). \end{aligned}$$

The force field is equal to zero in this case, however in the practical implementation we impose the boundary condition nodally and project the equation onto the nullspace of the (discrete) Dirichlet trace operator. In this manner, the boundary condition appears as a right hand side. Indeed, the weak form of (9) can be stated as: for any $\mu \in \mathbb{P}$, find $\hat{u}(\mu) \in H_0^1(\Omega)$ such that

$$a(\hat{u}(\mu), v; \mu) = -a(u_g, v; \mu), \quad \forall v \in H_0^1(\Omega),$$

where $u(\boldsymbol{\mu}) = \hat{u}(\boldsymbol{\mu}) + u_g$ and $u_g|_{\Gamma_g} = g$. Thus we have a modified linear form $\tilde{f}(v; \boldsymbol{\mu})$ with an affine decomposition inherited of the one from the bilinear form, i.e.

$$\tilde{f}(v; \boldsymbol{\mu}) = -a(u_g, v; \boldsymbol{\mu}) = -\sum_{q=1}^3 g_q(\boldsymbol{\mu}) a_q(u_g, v).$$

Figure 16 illustrates the local approximation spaces for some specific parameter values, the radius as a function of the parameter and the chosen sample points for $N = 20$ for the version with a fixed training set. We can observe that the symmetry of the problem is reflected in all three quantities and that the shape of the local approximation spaces does vary. Further, in Figure 16 we illustrate the training set, the local approximation spaces for some specific parameter values, the radius as a function of the parameter and the chosen sample points for $N = 20$ for the version with an adaptive training set.

In Figure 18 (left), we present the evolution of the error with respect to the number of computed truth solutions during the locally adaptive greedy algorithm for $N = 15, 20$ and also plot the convergence of the standard greedy algorithm, which of course converges faster but with a total number of basis functions equal to 121 at the on-line stage. In Figure 18 (right), the evolution of the accuracy on the training set of the version with the fixed and the adaptive training sets are compared. The value of about $4 \cdot 10^{-4}$ of the $L^\infty(\mathbb{P})$ -norm on the test grid is surprisingly inaccurate. Further numerical investigations have shown that the violation of the tolerance criterion is mainly focussed on the lower and upper boundary of the parameter space. The problem on the lower and upper boundary is caused how our implementation of generating the adaptive training set using FreeFem++ is placing the boundary points. Indeed, the criterion is violated on only 0.35% of all points of the test set. If one studies the error on a test grid of 87×99 points on $[-2.5, 0] \times [-0.7, 0.7]$ the error of the version with a fixed training set performs with an error of $1.13 \cdot 10^{-4}$ and the version with an adaptive training set shows an error of $1.06 \cdot 10^{-4}$.

We finally present in Table ?? some computation times of the on-line stage and compare it to a standard greedy-approach in order to highlight possible improvements that have to be made in a future work. As can be seen in Algorithm 4, the on-line stage consists of three steps: 1) Finding the N closest sample points, 2) effecting the on-line ortho-normalization and 3) solving the linear system. In Table ?? the average computing times over a sample of 10'000 random parameter values are given and the computations are done in Matlab restricting the computations to a single thread using a MacBook Pro with an Intel Core i7 processor with 2.66 GHz. We observe that the proposed locally adaptive greedy algorithm for $N = 20$ basis functions is significantly slower by about one order of magnitude than the standard greedy algorithm using $N = 121$ basis functions. While solving the linear system is of course faster for 20 unknowns than for 121, most time is spent by finding the closest sample points and by effecting the ortho-normalization. Certain improvements can certainly be done respecting the implementation of the proposed algorithm such as building a tree-like structure in the parameter space that indicates the considered sample points that might be used in each box. This tree can be constructed off-line. At the on-line stage and for a given parameter value $\boldsymbol{\mu}$ one has to identify the box that contains $\boldsymbol{\mu}$ and then searching for the N closest basis functions only among the sample points proposed by the selected box. Further, the triangular nature of the matrix γ in the on-line ortho-normalization process can be

exploited. In addition, the error estimation/certification is the most expensive part in most cases of the on-line stage of the reduced basis method. These computation times have not been accounted for since we worked with the exact error computation. Summarizing, these computation times indicate possible bottlenecks of the present approach but do not show the entire truth.

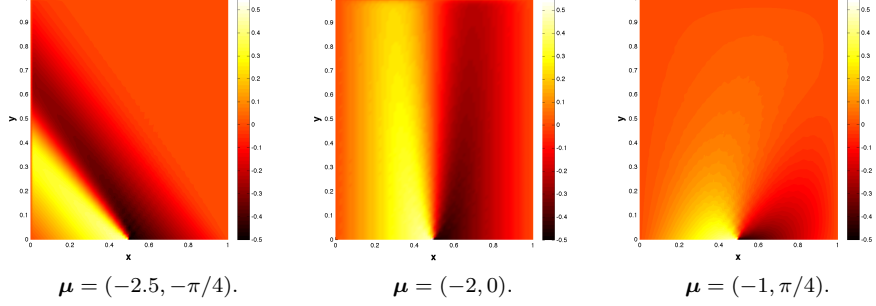


FIGURE 15. Test 4: Truth solutions for different parameter values which consist of the diffusion coefficient and the direction of the convection field.

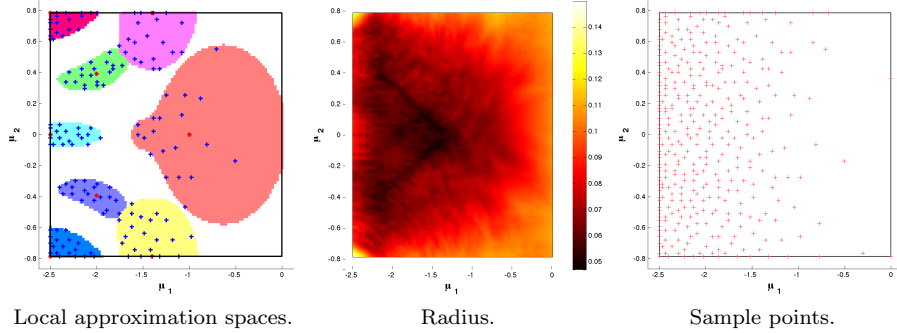


FIGURE 16. Test 4: Local approximation spaces for selected parameter values (left), radius as a function of the parameters (middle) and sample points (right) for $N = 20$ for the version using a fixed training set.

10. CONCLUSIONS

We presented a framework that can be seen as a generalization of the classical greedy algorithm that is widely used in reduced basis methods. The presented algorithm introduces local approximation spaces (in the parameter space) that also account for local anisotropic behavior instead of a global approach. The key idea was to consider the N closest basis functions (for a fixed N) where the distance is measured in an empirically built distance function which can be constructed on the fly and which is problem-dependent. In consequence, the number N of basis

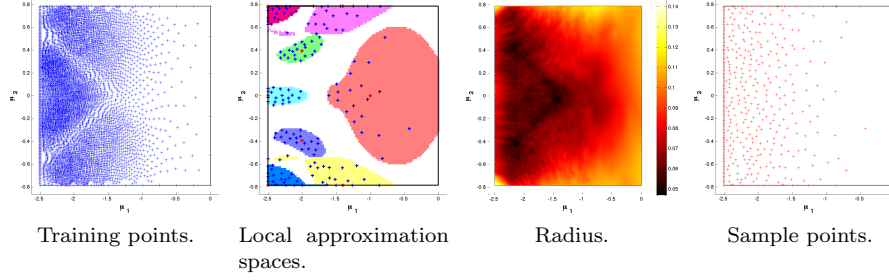


FIGURE 17. Test 4: Adaptive training set, local approximation spaces for selected parameter values, radius as a function of the parameters and sample points for $N = 20$ using the version using an adaptive training set.

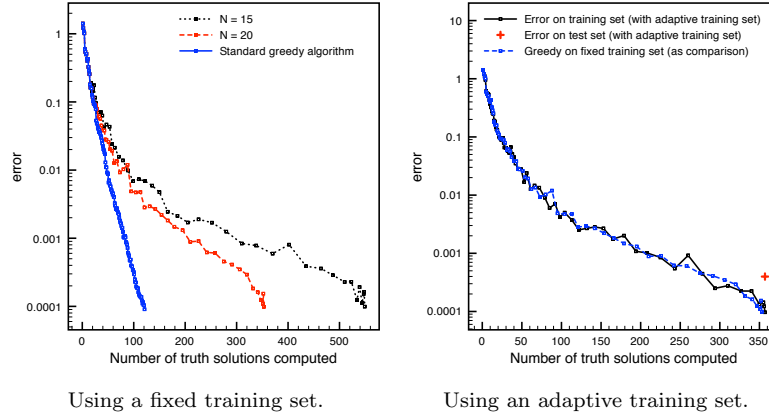


FIGURE 18. Test 4: Accuracy with respect to the number of computed truth solutions for the version with fixed training set (left) for $N = 15, 20$ and the version with an adaptive training set (right) for $N = 20$.

functions considered for an approximation is user controlled. Further, it is proposed to optionally place the training set uniformly distributed in this empirically constructed (approximative) metric space. Numerical tests illustrate the characteristics of this approach. In our future work, the emphasis will be shed on the implementation of this algorithm for high dimensional parameter spaces, the main challenge in this direction will be to design properly a set of equi-distributed points with an adapted non isotropic metric (in particular our current implementation).

ACKNOWLEDGEMENT

The authors want to thank the help and discussions with Frederic Hecht about the adaptation of FreeFem++ to their purpose. This work was also initiated out of the discussions during the workshop LJLL-SMP: "REDUCED BASIS METHODS

[ms]	Total on-line time	Search of N -closest sample points	Ortho-normalization	Solving the linear system
Fixed training set	2.9375	2.244	0.59266	0.10083
Adaptive training set	2.9182	2.2285	0.58874	0.10091
Standard greedy algorithm	0.3977	0	0	0.3977

TABLE 1. Test 4: Average on-line computation times over a random set of 10'000 parameter values.

IN HIGH DIMENSIONS” held in June 2011, organized with the support of the Fondation de Sciences Mathématiques de Paris.

REFERENCES

- [1] T. Bui-Thanh, K. Willcox, and O. Ghattas. Model reduction for large-scale systems with high-dimensional parametric input space. *SIAM J. Sci. Comput.*, 30(6):3270–3288, 2008.
- [2] C. D. Cantwell, S. J. Sherwin, R. M. Kirby, and P. H. J. Kelly. From h to p efficiently: strategy selection for operator evaluation on hexahedral and tetrahedral elements. *Comput. & Fluids*, 43:23–28, 2011.
- [3] I. Danaïla and F. Hecht. A finite element method with mesh adaptivity for computing vortex states in fast-rotating bose-einstein condensates. *J. Comput. Phys.*, 229(19):6946–6960, 2010.
- [4] J. L. Eftang, A. T. Patera, and E. M. Rønquist. An “hp” certified reduced basis method for parametrized elliptic partial differential equations. *SIAM J. Sci. Comput.*, 32(6):3170–3200, 2010.
- [5] J. L. Eftang and B. Stamm. Parameter multi-domain ‘hp’ empirical interpolation. *Internat. J. Numer. Methods Engrg.*, 90(4):412–428, 2012.
- [6] B. Fares, J. Hesthaven, Y. Maday, and B. Stamm. The reduced basis method for the electric field integral equation. *J. Comput. Phys.*, 230:5532–5555, 2011.
- [7] B. Haasdonk, M. Dihlmann, and M. Ohlberger. A training set and multiple bases generation approach for parameterized model reduction based on adaptive grids in parameter space. *Mathematical and Computer Modelling of Dynamical Systems*, 17(4):423–442, 2011.
- [8] B. Haasdonk and M. Ohlberger. Reduced basis method for finite volume approximations of parametrized linear evolution equations. *M2AN Math. Model. Numer. Anal.*, 42(2):277–302, 2008.
- [9] F. Hecht, O. Pironneau, A. L. Hyaric, and K. Ohtsuka. *FreeFEM++ Manual*, 2008.
- [10] P. Holmes, J. L. Lumley, and G. Berkooz. *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge Monographs on Mechanics. Cambridge University Press, Cambridge, 1996.
- [11] B. H. M. Dihlmann, S. Kaulmann. Online reduced basis construction procedure for model reduction of parametrized evolution systems. Technical report, University of Stuttgart, SimTech Preprint 2012-1, 2012.
- [12] N.-C. Nguyen, G. Rozza, and A. T. Patera. Reduced basis approximation and a posteriori error estimation for the time-dependent viscous Burgers’ equation. *Calcolo*, 46(3):157–185, 2009.
- [13] A. T. Patera and G. Rozza. *Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Partial Differential Equations*. 2006.
- [14] C. Prud’homme, D. V. Rovas, K. Veroy, Y. Maday, A. T. Patera, and G. Turinici. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *Journal of Fluids Engineering*, 124(1):70–80, Mar 2002.

- [15] A. Quarteroni, G. Rozza, and A. Manzoni. Certified reduced basis approximation for parametrized partial differential equations and applications. *Journal of Mathematics in Industry*, 1(1):3–3, Jun 2011.
- [16] G. Rozza, D. B. P. Huynh, and A. T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations: application to transport and continuum mechanics. *Arch. Comput. Methods Eng.*, 15(3):229–275, 2008. on.
- [17] S. Sen, K. Veroy, D. B. P. Huynh, S. Deparis, N. C. Nguyen, and A. T. Patera. “Natural norm” a posteriori error estimators for reduced basis approximations. *J. Comput. Phys.*, 217(1):37–62, 2006.
- [18] K. Veroy, C. Prud’homme, D. V. Rovas, and A. T. Patera. A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations (AIAA paper 2003-3847). *Proceedings*, Jun 2003.

UPMC UNIVERSITY PARIS 06, UMR 7598, LABORATOIRE JACQUES-LOUIS LIONS, F-75005, PARIS, FRANCE, AND DIVISION OF APPLIED MATHEMATICS, BROWN UNIVERSITY, PROVIDENCE RI, USA.

E-mail address: `maday@ann.jussieu.fr`

UPMC UNIVERSITY PARIS 06, UMR 7598, LABORATOIRE JACQUES-LOUIS LIONS, F-75005, PARIS, FRANCE AND CNRS, UMR 7598, LABORATOIRE JACQUES-LOUIS LIONS, F-75005, PARIS, FRANCE.

E-mail address: `stamm@ann.jussieu.fr`